Assignment 2: AMS 268 (Due Date 2/28)

February 14, 2016

(a) Consider the data $(y_i, \boldsymbol{x}_i)_{i=1}^n$ is following a high dimensional probit regression model

$$P(y_i = 1 | \boldsymbol{x}_i) = \Phi(\boldsymbol{x}_i' \boldsymbol{\beta}),$$

where $y_i \in \{0, 1\}$, $\boldsymbol{x}_i \in \mathcal{R}^p$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$. Consider $p = 400$, $n = 1000$.

(i) Simulate 5 datasets by drawing $\boldsymbol{x}_i \sim N(\boldsymbol{0}, \boldsymbol{I})$ and $\boldsymbol{\beta} = (1.3, 4, -1, 1.6, 5, -2, \boldsymbol{0}_{394 \times 1})'$.

(ii) Compare parameter estimates from MCMC and Sequential Monte Carlo (SMC).

(iii) Provide the coverage and length of 95% credible intervals of $\beta_1, ..., \beta_6$ for MCMC and SMC.

(b) Suppose

$$y_i = x_{1i} + (x_{2i} - 0.5)_+ + x_{3i}^2 + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2),$$

for $i = 1, ..., 5000$. Consider, $\sigma^2 = 0.5$ and $x_{1i}, x_{2i}, x_{3i} \sim N(0, 1)$.

(i) Use the above model to generate 5 datasets.

(ii) Run a predictive process model with the above data by writing your own code.

(iii) Comment on the model fit in terms of estimating the mean function and error variance.